

CONTINUATION-IN-PART APPLICATION

UNDER 37 CFR § 1.53(B)

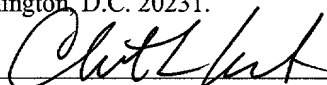
TITLE: **EXCEPTION HANDLING METHOD AND  
APPARATUS FOR USE IN PROGRAM CODE  
CONVERSION**

APPLICANT(S): Rawsthorne, Alasdair; Sandham, John H.;  
Souloglou, Jason

Documents Enclosed:

Specification (18 pgs); Claims (6 pgs);  
Abstract (1pg); Related Applications (1 pg);  
Drawings (3 pgs); Declaration (2 pgs); Assignment  
(2 pgs); Recordation Cover Sheet (3pgs); Small  
Entity Statement (2 pg); Check in the Amt. of  
\$515.00

"EXPRESS MAIL" Mailing Label Number EL573655657US Date of Deposit April 6, 2001  
I hereby certify under 37 CFR §1.10 that this correspondence is being deposited with the  
United States Postal Service as "Express Mail Post Office to Addressee" with sufficient  
postage on the date indicated above and is addressed to the Commissioner for Patents,  
Washington, D.C. 20231.

  
\_\_\_\_\_  
Christie L. Martin

**EXCEPTION HANDLING METHOD AND APPARATUS FOR USE  
IN PROGRAM CODE CONVERSION**

5       The present invention relates in general to the field of program code conversion for computer systems and in particular, but not exclusively, to the field of emulation using dynamic binary translation.

10       A program for a computer takes several different forms. Typically a program is written first in a high-level language readily understood by a human programmer. The program is compiled from the high-level language into a low-level language more appropriate for control of a  
15       computer's processor and related components. However, in order for the processor to function, the program code must be provided in a machine-readable form that directs primitive operations of the processor such as loading, shifting, adding and storing operations. To run faster,  
20       some processors use an expanded set of instructions each of which represent a sequence of primitive instructions. This is known as a complicated instruction set computer (CISC). However, program code written (or compiled) specifically for a first processor with a particular  
25       instruction set in most cases cannot run on any other type of processor because of differences between the instruction set for each type of processor.

      An emulator allows program code written for a  
30       processor of a first type (a "subject" processor) to be run on a processor of a second type (a "target" or "host" processor). One form of this emulation process is known as binary translation because executable binary code

appropriate to the subject processor is translated into executable binary code appropriate to the target processor.

5        In static binary translation an entire program is translated prior to execution of the translated program on the target processor. This involves a significant delay. Therefore, emulators have been developed which employ dynamic binary translation to translate small sections of  
10 a source program for execution immediately on the target processor. This is much more efficient because large sections of the source code will not be used in practice, or will be used only rarely. An emulator employing a dynamic translation system selects only the required parts  
15 of the source program for translation on demand, as the program is run.

There now follows a brief summary of dynamic binary translation for an emulator as may be employed in a  
20 preferred embodiment of the present invention. More detailed background in the field of dynamic binary translation is given in the applicant's co-pending application number GB 98 22075.9 entitled "PROGRAM CODE CONVERSION" the content of which is incorporated herein by  
25 reference to avoid wasteful duplication.

When performing dynamic binary translation the emulator appears to the subject code as if the subject code were running on the appropriate subject processor.  
30 The emulator replicates the subject machine including, for example, registers of the subject processor, such that the emulator provides a virtual subject machine. The emulated registers are termed herein "abstract registers" and

correspond to the set of registers of the subject processor used by the subject code. In the preferred dynamic translation process the emulator first translates a predetermined small section of the subject code into an  
5 intermediate representation which represents the instructions of the subject code in a generic format, optionally performs optimisation on the intermediate representation, and then translates the optimised intermediate representation into executable binary code  
10 for the target processor. It is preferred that the small section of subject code corresponds to a "basic block" which starts with a first instruction at a unique entry point and ends at a last instruction at an unique exit point. Typically the last instruction of the block is a  
15 jump, call or branch instruction (conditional or unconditional).

In the field of binary translation, a problem arises with respect to the handling of exceptions. An exception  
20 indicates that a condition has occurred which needs to be handled before processing can continue. This includes explicit exceptions performed as an instruction in the subject code (for example an exception is reported if the value of one register is greater than the value of a  
25 second register), and implicit exceptions which occur for example as a result of memory read or write operations to a memory page that is not currently available. In both cases the exception is desirably reported to an exception handler written in subject code. However, many subject  
30 machine architecture definitions require that the exception is reported on a boundary between subject code instructions, following a predetermined set of rules. For example, the subject architecture definition may require

that when the exception is reported, the effects of all previous subject instructions are complete, the exception points to the first instruction of the subject code which has not been executed and no effects from that subject instruction or any subsequent instruction have yet taken place. Further, the architecture definition of a particular processor may have different rules for different types of exceptions.

10 In the context of binary translation it is apparent that a target instruction performed on the target processor that causes an exception to be reported will not of itself fulfil the conditions for reporting the exception to an exception handler written in subject code.

15 Instructions are almost always performed on the target processor in a different order compared with the order of instructions in the corresponding block of subject code, firstly due to the differences between the instruction set of the subject processor for which the subject code was

20 written and the target processor on which the translated target code is run, and secondly because of the optimisation of the intermediate representation that typically occurs during translation.

25 Exceptions can occur in response to execution of the translated target code on the target processor, and can occur during execution of the emulator code on the target processor, i.e. during translation. In order to report an exception to the subject exception handler the state of

30 the virtual subject processor represented by the emulator must be available to the subject exception handler, including the correct status of the registers of the subject processor.

One approach to this problem is to return the virtual subject machine to the conditions that applied at entry into the section of code being translated or executed, i.e. by returning the virtual subject machine to the condition prevailing at the point of entry into the current block of subject code instructions being translated or executed. The exception handler can now step through the instructions of the block of source code individually in sequence until the instruction causing the exception is identified.

US 5832205 (Kelly et al) discloses an emulator which uses a set of "working" registers during emulation of each section of subject code. The content of each of these working registers is copied to a set of "official" virtual subject registers at the end of the section of subject code, using a gated store buffer. Therefore, if an exception occurs during emulation of a section of subject code this will affect only the working registers and the condition of the virtual subject machine can be recovered from the "official" registers at the point of entry into that section of subject code. However, the use of "working" and "official" registers adds significantly to the overhead of the emulation process in the target processor due to the copying of information from the "working" registers to the corresponding "official" registers at the end of each section of subject code.

An aim of the present invention is to provide a method of representing subject registers in an emulator which allows exceptions to be accurately reported to a subject exception handler in accordance with the rules of the

handler, whilst minimising overhead in the emulator. It is a further aim of the present invention to provide an emulation method and apparatus wherein subject registers are represented to allow accurate exception handling.

5

According to the present invention there is provided a method of representing a subject register in an emulator, the method comprising the steps of:

- 10 (a) mapping an abstract register representing a subject register of a subject machine to either a first location or to a second location within a target machine; and
- 15 (b) alternating mapping of the abstract register between the first and second locations such that the content of one of the first or second locations represents a definitive version of the abstract register for use by the emulator during exception handling, whilst the
- 20 other of the first or second locations represents a speculative version of the abstract register.

Preferably, for a predetermined section of subject code, one of the first or second locations holds a definitive value of the abstract register at entry into that section, whilst the other of the first or second locations holds a speculative current value of the abstract register for use during that section.

30 Preferably, the abstract register is mapped to the alternate one of the first or second locations upon reaching the end of the section of subject code. The speculative version becomes definitive when it is

determined that no exception has occurred in the section of subject code. Ideally, the step of alternating mapping is performed only if the content of the speculative version of the abstract register has been updated during  
5 the predetermined section of subject code.

Preferably, the step (a) comprises the steps of: (a1) providing a plurality of abstract registers each representing a register of the subject machine; (a2)  
10 mapping each of the plurality of abstract registers to either a respective one of a first set of locations or a respective one of a second set of locations within the target machine. Preferably, the step (b) comprises alternating mapping for each of the abstract registers  
15 between the respective one of each of the first and second sets of locations.

Preferably, the first location and/or the second location is a memory location within the target machine.  
20 Alternatively, the first location and/or the second location is a register of the target machine.

Preferably, the method is for use with an emulator that performs dynamic binary translation. Suitably, the  
25 predetermined section of subject code represents one or more basic blocks of subject code.

Also according to the present invention there is provided a method for use in handling exceptions by an  
30 emulator performing program code conversion between subject code suitable for a subject processor and target code suitable for a target processor, the method comprising the steps of providing at least one abstract



register (X,Y) each representing a register of the subject processor; (b) mapping the or each abstract register to a corresponding pair of locations within the target processor; and (c) alternating mapping of the or each  
5 abstract register between a first of the pair of locations and a second of the pair of locations, such that for a predetermined section of subject code, one of the first or second locations holds a definitive value of the abstract register at entry into that section for use by the  
10 emulator during exception handling, whilst the other of the first or second locations holds a speculative current value of the abstract register for updating by the emulator during that section.

15 According to further aspects of the present invention there is provided an emulator method and an emulator apparatus for performing the method according to any statement herein. The present invention also extends to a computer when programmed to perform the method according  
20 to any statement herein, to a computer program for performing the method according to any statement herein, and to a computer program product containing computer readable instructions for performing the method according to any statement herein.

25

For a better understanding of the invention, and to show how embodiments of the same may be carried into effect, reference will now be made, by way of example, to the accompanying diagrammatic drawings, in which:

30

Figure 1 shows a typical prior art configuration for a subject processor;

Figure 2 shows a typical emulator using binary translation;

Figure 3 shows a configuration of an emulator using  
5 binary translation as may be employed in preferred  
embodiments of the present invention;

Figure 4 shows a preferred binary translation type  
emulator in use; and

10

Figures 5 and 6 show example sets of abstract  
registers.

Referring to Figure 1 a typical prior art arrangement  
15 is shown illustrating the configuration of a subject  
machine wherein subject code 10 is executed directly on a  
subject processor 11. Suitably the subject code is  
executable binary code. However, the subject code may be  
represented in any suitable language with intermediate  
20 layers (compilers, etc.) between the subject code 10 and  
the subject processor 11 as will be familiar to the  
skilled person.

Referring now to Figure 2, a typical prior art  
25 configuration is shown to illustrate the use of a binary  
translation type emulator 20 as an intermediate layer  
enabling the subject code 10 to be executed by a target  
processor 31. The preferred embodiment of the present  
invention is particularly intended for use with an  
30 emulator 20 which performs dynamic binary translation of  
the subject code 10 into target code 30 executable on the  
target processor 31.

Referring now to Figure 3 the emulator 20 of the preferred embodiment is illustrated in more detail and comprises a front end 21, a core 22 and a back end 23.

5       The front end 21 is configured specific to the subject processor 11 being emulated. The front end 21 translates instructions of the subject code 10 into a generic intermediate representation for each basic block of subject code. Each basic block suitably includes a  
10 sequential set of instructions between a first instruction representing a unique entry point and a last instruction at a unique exit point (such as a jump, call or branch instruction). In a particularly preferred embodiment the emulator 20 selects a group block comprising two or more  
15 basic blocks chosen for code generation and optimisation as a single unit. Further, the emulator 20 supports iso-blocks representing the same basic block of subject code under different entry conditions. Each predetermined section of the subject code 10 results in a block of  
20 intermediate representation (an "IR block").

The core 22 optimises each IR block generated by the front end 21 by employing optimisation techniques which need not be described here in detail. The back end 23  
25 takes optimised IR blocks from the core 22 and produces target code 30 executable by the target processor 31.

As shown in Figure 4, in use a first predetermined section of the subject code 10 is identified such as a  
30 basic block 100 and translated by the emulator 20 running on the target processor 31 in a translation mode. The target processor 31 then executes the corresponding optimised and translated block 300 of target code 30.

The preferred emulator 20 includes a plurality of abstract registers, suitably provided in the core 22 shown in Figure 3, which represent the physical registers that are used within the subject processor 11 to execute the subject code 10. The abstract registers define the state of the subject processor 11 being emulated by representing the expected effects of the subject code instructions on the registers of the subject processor.

10

As shown in Figure 5, in the preferred embodiment two sets of abstract registers are provided, here labelled set A and set B. At initialisation a first of these two sets, for example set A, holds "definitive" values. That is, the registers of set A are defined to hold initial values which are known to be valid representing the expected content of the physical registers of the subject processor 11 being emulated.

The second set of abstract registers, in this example labelled set B, are initially defined to represent "speculative" values. That is, at initialisation the second set of abstract registers (set B) also hold the expected initial content of the physical registers of the subject processor 11, but the values in set B are not relied upon as being valid.

In the translation process the emulator 20 uses the speculative set of abstract registers (i.e., set B) such that the content is updated to show the expected state of the physical registers of the subject processor 11 after execution of the block 100 of subject code 10. During translation of the first block, the content of the

definitive set of abstract registers of (i.e., set A) remains unchanged.

If an exception occurs during translation or execution  
5 of the basic block 100, then the emulator 20 readily  
recovers the condition of the subject registers upon entry  
into the block 100 using the abstract registers marked as  
holding definitive data (i.e., set A). The exception  
handler can now step through the instructions of the block  
10 100 of the source code 10 in sequence until the  
instruction causing the exception is identified. Here,  
the status of the abstract registers is updated after each  
instruction. Therefore, when the subject code instruction  
responsible for the exception is identified the condition  
15 of the virtual subject machine represented by the emulator  
are reported to the subject code exception handler  
according to the rules thereof. The subject code  
exception handler recovers the exception in accordance  
with the handling process and returns to a point in the  
20 subject code appropriate to the exception. For example, it  
is common that the exception handler returns the emulator  
to the next unexecuted instruction of the source code 10.  
The translation or execution process can then continue  
from that point.

25

After the successful execution of the translated block  
of target code 300 the registers in the set holding  
speculative values (i.e., set B) will have been updated to  
hold the expected content of the equivalent registers of  
30 the subject processor 11 being emulated at the end of the  
basic block 100 of subject code 10. At this point, the  
abstract registers of the first set (i.e., set A) hold the  
condition of the subject processor at entry to the block

of subject code 100, and the abstract registers of the second set (i.e., set B) hold the condition at the end of the block 100. Since the block 100 has been successfully translated and executed, the abstract registers of set B  
5 are now defined as holding definitive values, and the abstract registers of set A are defined as holding speculative values.

The abstract registers of set A and set B suitably  
10 form register pairs. Each register in set A has a corresponding partner register in set B. One of the pair holds the definitive value of that abstract register, whilst the other holds the speculative value. At the end of each section of code the definition of these two  
15 registers is reversed such that each register of the pair performs the opposite function during the next section of code. Alternating the function of the two registers of each register pair provides a simple and elegant method of maintaining the entry conditions for the current section  
20 of code.

As a further advantage, it is not necessary to update the status of every abstract register upon successful completion of each section of code. Only those registers  
25 which have been changed during that section need be updated to show their new function. If the value of a particular abstract register is not changed during a section of code then that value will remain in place during the next section to perform the same definitive or  
30 speculative function as appropriate.

Referring now to Figures 5 and 6, a simplified example embodiment will now be described. At step 1 shown in

Figure 5, a first abstract register (Reg  $X_A$ ) representing register X of the subject processor 11 contains the definitive value whilst a second abstract register (Reg  $X_B$ ) contains the speculative value. The abstract registers are suitably held in memory locations and a working map for register X points to the location of the speculative version Reg  $X_B$ . Similarly, for register Y in this example initially Reg  $Y_A$  is definitive whilst Reg  $Y_B$  is speculative. After performance of one or more instructions, such as the block 100 of the subject code 10, the mapping for the abstract registers is updated for step 2 as shown in Figure 6. In this example, the content of the speculative register Y has changed and therefore Reg  $Y_B$  is now taken to be the definitive version for use in a subsequent block. The map for register Y is updated to point to Reg  $Y_A$  as the speculative version. By contrast, register X was not affected by the instruction or instructions in the block 100 and therefore Reg  $X_A$  remains as the definitive version whilst Reg  $X_B$  remains as the speculative version.

To allow continuity of register content between sections of code, suitably the first read operation encountered during a current section of code uses the definitive version of each particular abstract register. The definitive version represents the condition of that register at entry into the current section of code and therefore maintains continuity with the previous section. Further read operations also use the definitive version of each abstract register, until a write operation is encountered. The first write operation uses the speculative version of each particular abstract register. Therefore the definitive version remains unchanged and the

speculative version now contains the current value of the relevant abstract register. Subsequent read and write operations use the speculative version for the remainder of that section of code.

5

As described above, in preferred embodiments of the present invention an abstract register are provided corresponding to each physical register of the subject processor 11, with the abstract register being mapped to  
10 two predetermined locations. One of each pair of abstract register locations contains a definitive value whilst the other contains a speculative value. The function of these two locations is readily reversed to alternate the location holding definitive content. Therefore, time  
15 consuming copying operations are avoided.

In the preferred embodiment, two versions of each of abstract register are achieved by mapping to two sets of memory locations and the definitive and speculative  
20 versions alternated by alternating the memory mapping between these two locations. Updating the map of the abstract registers held in the target machine to replicate the physical registers of the subject processor is performed quickly and simply at translation time, and  
25 imposes no overhead when the translated code is executed, possibly many times.

In a further preferred embodiment, one or both of the definitive and speculative versions of the abstract  
30 registers may be stored in a pair of target machine registers (on a target machine with a sufficiency thereof) as an alternative to using a pair of memory locations.



One aspect of the preferred embodiment of the present invention addresses the problem of fixing register references across branches, and in particular a branch (or loop) from a current block of subject code to a previously translated block. Branches between blocks of code are commonly encountered in practice, and often involve the same block of code being referenced from more than one other locations within the subject code, thereby generating different entry conditions. One solution is to copy the content from the definitive to the speculative version of the abstract register, such that the entry conditions are appropriate for use of the previously translated block of code. For example, when the block was first translated the first version Reg  $X_A$  was definitive. Hence, if the same block of translated code is to be used again subsequently, and it is reached in a condition where the second version of the abstract register Reg  $X_B$  is definitive, then the content of register  $X_B$  must be copied to register Reg  $X_A$  before continuing. This incurs a copying overhead when implementing a branch to a previously translated block of code. A preferred solution employed in embodiments of the present invention avoids this copying operation by translating the block again, this time under the conditions prevailing at the time of the branch, such that plural versions of the previously translated code now exist each associated with a particular set of entry conditions. Although increased translation work is involved, copying overhead during execution is avoided. In general, translating a section of subject code (containing any number of basic blocks) by one extra time eliminates the compensation copying that would have been necessary for any particular register that was updated an odd number of times in that loop.

Although the embodiments described above refer to an emulator employing dynamic binary translation, the method is also applicable to static translation where a large section of code is translated prior to execution. In static translation the section of code selected for translation typically represents a whole program or a major part of a program. However, it is still convenient to use the method described above for handling exceptions arising during translation and execution of the translated code, enabling exception handling to be performed at least from the condition at entry into that section of code. Further, the method is applicable to program code optimisation wherein the subject machine and the target machine have the same or at least compatible instruction sets and architectures.

The present invention extends to a computer when programmed to perform the method described above, to a computer program for performing the method described above, and to a computer program product containing computer readable instructions for performing the method described above.

The reader's attention is directed to all papers and documents which are filed concurrently with or previous to this specification in connection with this application and which are open to public inspection with this specification, and the contents of all such papers and documents are incorporated herein by reference.

All of the features disclosed in this specification (including any accompanying claims, abstract and

drawings), and/or all of the steps of any method or process so disclosed, may be combined in any combination, except combinations where at least some of such features and/or steps are mutually exclusive.

5

Each feature disclosed in this specification (including any accompanying claims, abstract and drawings), may be replaced by alternative features serving the same, equivalent or similar purpose, unless expressly stated otherwise. Thus, unless expressly stated otherwise, each feature disclosed is one example only of a generic series of equivalent or similar features.

The invention is not restricted to the details of the foregoing embodiment(s). The invention extend to any novel one, or any novel combination, of the features disclosed in this specification (including any accompanying claims, abstract and drawings), or to any novel one, or any novel combination, of the steps of any method or process so disclosed.